



430

*Inventors*

PATENT

Attorney Docket No. 3495.0010-01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

#30  
(283)

In re Application of:

Marc Alizon, et al.

Serial No.: 07/158,652

Filed: February 22, 1988

Group Art Unit: 18054

Examiner: J. Railey

For: CLONED DNA SEQUENCES RELATED  
TO THE GENOMIC RNA OF  
LYMPHADENOPATHY ASSOCIATED  
VIRUS (LAV) AND PROTEINS  
ENCODED BY SAID LAV GENOMIC  
RNA

Honorable Commissioner of Patents  
and Trademarks  
Washington, D.C. 20231

Sir:

CLAIM FOR PRIORITY

Under the provisions of Section 119 of 35 U.S.C.,  
applicants hereby claim the benefit of the filing date of Great  
Britain Application No. 84 29099, filed November 16, 1984, for  
the above identified United States Patent Application.

In support of applicants' claim for priority, filed  
herewith is one certified copy of GB 84 29099.

If there are any fees due in connection with the filing of  
this Paper, please charge such fees to our Deposit Account  
No. 06-0916.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,  
GARRETT & DUNNER

By: Michele M. Schafer  
Michele M. Schafer  
Reg. No. 34,717

Date: October 21, 1993



THE PATENT OFFICE,  
25 SOUTHAMPTON BUILDINGS,  
LONDON.

I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents, has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Please turn over



Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

21<sup>st</sup> Witness my hand this  
day of OCTOBER 19 85

COC.1

*M. Russell*

# PATENTS ACT 1977

16 NO 384

PATENTS FORM No. 1/77 (Revised 1982)

(Rules 16, 19)

The Comptroller  
The Patent Office  
25 Southampton Buildings  
London, WC2A 1AY

19/11/84 B3662 PAT\*\*\* 10.00

1984

29099

Fee: £10.00

## REQUEST FOR GRANT OF A PATENT

8429099

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

I Agent's Reference JJD/EPF/26804

II Title of Invention CLONED DNA SEQUENCES RELATED TO THE GENOMIC RNA OF LYMPHADENOPATHY-ASSOCIATED VIRUS (LAV) AND PROTEINS ENCODED BY SAID LAV GENOMIC RNA.

III Applicant or Applicants (See note 2)

Name (First or only applicant) INSTITUT PASTEUR

Country FRANCE

State

ADP Code No.

Address 25-28 Rue du Dr. Roux,

75724 Paris Cedex 15, France.

Name (of second applicant, if more than one) Centre National

de la Recherche Scientifique

Country FRANCE

State

Address 15 Quay Anatole France

75007 Paris, France.

IV Inventor (see note 3)

~~Not applicable to this form~~

or

(b) A statement on Patents Form No. 7/77 is/will be furnished

V Name of Agent (if any) (See note 4)

**Reddie & Grose**

ADP CODE NO

VI Address for Service (See note 5)

**16 Theobalds Road**

**London WC1X 8PL**

VII Declaration of Priority (See note 6)

Country

Filing date

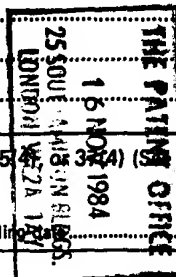
File number

VIII The Application claims an earlier date under Section 8(3), 12(6), 15(4), 16(3) (See note 7)

Section No.

Earlier application or patent number

and filing



**IX Check List (To be filled in by applicant or agent)**

- |  |  |
|--|--|
| <b>A</b> The application contains the following number of sheet(s) | <b>B</b> The application as filed is accompanied by:-                |
| 1 Request ..... <u>1</u> ..... Sheet(s)                            | 1 Priority document ..... <u>No</u> .....                            |
| 2 Description ..... <u>17</u> ..... Sheet(s)                       | 2 Translation of priority document ..... <u>No</u> .....             |
| 3 Claim(s) ..... <u>2</u> ..... Sheet(s)                           | 3 Request for Search ..... <u>No</u> .....                           |
| 4 Drawing(s) ..... <u>26</u> ..... Sheet(s)                        | 4 Statement of Inventorship and Right to Apply ..... <u>No</u> ..... |
| 5 Abstract ..... <u>0</u> ..... Sheet(s)                           | 5  |

**X** It is suggested that Figure No ..... 1 ..... of the drawings (if any) should accompany the abstract when published.

**XI** Signature (See note 8)

  
Reddie & Grose, Agents for the Applicant(s)

**NOTES:**

1. This form, when completed, should be brought or sent to the Patent Office together with the prescribed fee and two copies of the description of the invention, and of any drawings. ✓
2. Enter the name and address of each applicant. Names of individuals should be indicated in full and the surname or family name should be underlined. The names of all partners in a firm must be given in full. Bodies corporate should be designated by their corporate name and the country of incorporation and, where appropriate, the state of incorporation within that country should be entered where provided. Full corporate details, eg "a corporation organised and existing under the laws of the State of Delaware, United States of America," trading styles, eg "trading as xyz company", nationality, and former names, eg "formerly [known as] ABC Ltd." are not required and should not be given. Also enter applicant(s) ADP Code No. (if known).
3. Where the applicant or applicants is/are the sole inventor or the joint inventors, the declaration (a) to that effect at IV should be completed, and the alternative statement (b) deleted. If, however, this is not the case the declaration (a) should be struck out and a statement will then be required to be filed upon Patent Form No 7/77.
4. If the applicant has appointed an agent to act on his behalf, the agent's name and the address of his place of business should be indicated in the spaces available at V and VI. Also insert agent's ADP Code No. (if known) in the box provided.
5. An address for service in the United Kingdom to which all documents may be sent must be stated at VI. It is recommended that a telephone number be provided if an agent is not appointed.
6. The declaration of priority at VII should state the date of the previous filing and the country in which it was made and indicate the file number, if available.
7. When an application is made by virtue of section 8(3), 12(6), 15(4), or 37(4) the appropriate section should be identified at VIII and the number of the earlier application or any patent granted thereon identified.
8. Attention is directed to rules 90 and 106 of the Patent Rules 1982.
9. Attention of applicants is drawn to the desirability of avoiding publication of inventions relating to any article, material or device intended or adapted for use in war (Official Secrets Acts, 1911 and 1920). In addition after an application for a patent has been filed at the Patent Office the comptroller will consider whether publication or communication of the invention should be prohibited or restricted under section 22 of the Act and will inform the applicant if such prohibition is necessary.
10. Applicants resident in the United Kingdom are also reminded that, under the provisions of section 23 applications may not be filed abroad without written permission or unless an application has been filed not less than six weeks previously in the United Kingdom for a patent for the same invention and no direction prohibiting publication or communication has been given or any such direction has been received.

Cloned DNA sequences related to the genomic RNA of lymphadenopathy-associated-virus (LAV) and proteins encoded by said LAV genomic RNA

5 The invention relates to cloned DNA sequences indistinguishable from genomic RNA and DNA of lymphadenopathy-associated virus (LAV), a process for their preparation and their uses. It relates more particularly to stable probes including a DNA sequence which can be used for the detection of the LAV virus or related viruses  
10 or DNA proviruses in any medium, particularly biological samples containing any of them. The invention also relates to polypeptides, whether glycosylated or not, encoded by said DNA sequences.

Lymphadenopathy-associated virus (LAV) is a human  
15 retrovirus first isolated from the lymph node of a homosexual patient with lymphadenopathy syndrome, frequently a prodrome or a benign form of acquired immune deficiency syndrome (AIDS). Subsequently other LAV isolates have been recovered from patients with AIDS or pre-AIDS. All available data are consistent with the virus being the causative  
20 agent of AIDS.

A method for cloning such DNA sequences has already been disclosed in British Patent Application Nr. 84 23659 filed on September 19, 1984. Reference is here-  
25 after made to that application as concerns subject matter in common with the further improvements to the invention disclosed herein.

The present invention aims at providing additional new means which should not only also be useful for the  
30 detection of LAV or related viruses (hereafter more generally referred to as "LAV viruses"), but also have more versatility, particularly in detecting specific parts of the genomic DNA of said viruses whose expression products are not always directly detectable by immunological  
35 methods.

The present invention further aims at providing

polypeptides containing sequences in common with polypeptides encoded by the LAV genomic RNA. It relates even more particularly to polypeptides comprising antigenic determinants included in the proteins encoded and expressed by the LAV genome occurring in nature. An additional object of the invention is to further provide means for the detection of proteins related to LAV virus, particularly for the diagnosis of AIDS or pre-AIDS or, to the contrary, for the detection of antibodies against the LAV virus or proteins related therewith, particularly in patients afflicted with AIDS or pre-AIDS or more generally in asymptomatic carriers and in blood-related products. Finally the invention also aims at providing immunogenic polypeptides, and more particularly protective polypeptides for use in the preparation of vaccine compositions against AIDS or related syndromes.

The present invention relates to additional DNA fragments, hybridizable with the genomic RNA of LAV as they will be disclosed hereafter, as well as with additional cDNA variants corresponding to the whole genomes of LAV viruses. It further relates to DNA recombinants containing said DNAs or cDNA fragments.

The invention relates more particularly to a cDNA variant corresponding to the whole of LAV retroviral genomes, which is characterized by a series of restriction sites in the order hereafter (from the 5' end to the 3' end).

The coordinates of the successive sites of the whole LAV genome (restriction map) are indicated hereafter too, with respect to the Hind III site (selected as of coordinate 1) which is located in the R region. The coordinates are estimated with an accuracy of  $\pm 200$  bp :

	Hind III	0
	Sac I	50
35	Hind III	520
	Pst I	800
	Hind III	1 100

	Bgl II	1 500
	Kpn I	3 500
	Kpn I	3 900
	Eco RI	4 100
5	Eco RI	5 300
	Sal I	5 500
	Kpn I	6 100
	Bgl II	6 500
	Bgl II	7 600
10	Hind III	7 850
	Bam HI	8 150
	Xho I	8 600
	Kpn I	8 700
	Bgl II	8 750
15	Bgl II	9 150
	Sac I	9 200
	Hind III	9 250

Another DNA variant according to this invention optionally contains an additional Hind III approximately at the 5 550 coordinate.

Reference is further made to fig. 1 which shows a more detailed restriction map of said whole-DNA (AJ19).

An even more detailed nucleotide sequence of a preferred DNA according to the invention is shown in fig. 4-12 hereafter.

The invention further relates to other preferred DNA fragments which will be referred to hereafter.

Additional features of the invention will appear in the course of the non-limitative disclosure of additional features of preferred DNAs of the invention, as well as of preferred polypeptides according to the invention. Reference will further be had to the drawings in which :

- fig. 1 is the restriction map of a complete LAV genome (clone AJ19) ;
- figs. 2 and 3 show diagrammatically parts of the three

4

possible reading phases of LAV genomic RNA, including the open reading frames (ORF) apparent in each of said reading phases :

- figs. 4-12 show the successive nucleotidic sequences of a complete LAV genome. The possible peptide sequences in relation to the three possible reading phases related to the nucleotidic sequences shown are also indicated ;
- figs. 13-18 reiterate the sequence of part of the LAV genome containing the genes coding for the envelope proteins, with particular boxed peptidic sequences which corresponds to groups which normally carry glycosyl groups.

The sequencing and determination of sites of particular interest was carried out on a phage recombinant corresponding to AJ19 disclosed in the abovesaid British Patent application Nr. 84 23659. A method for preparing it is disclosed in that application.

The whole recombinant phage DNA of clone AJ19 (disclosed in the earlier application) was sonicated according to the protocol of DEININGER (1983), Analytical Biochem. 129, 216. the DNA was repaired by a Klenow reaction for 12 hours at 16°C. The DNA was electrophoresed through 0.8 % agarose gel and DNA in the size range of 300-600 bp was cut out and electroeluted and precipitated. Resuspended DNA (in 10 mM Tris, pH 8 ; 0.1 mM EDTA) was ligated into M13mp8 RF DNA (cut by the restriction enzyme SmaI and subsequently alkaline phosphated), using T4 DNA- and RNA-ligases (Maniatis T et al (1982) - Molecular cloning - Cold Spring Harbor Laboratory). An *E. coli* strain designated as TGI was used for further study. This strain has the following genotype :

*Alac pro, supE, thi.F'treD36, proAB, lacI<sup>q</sup>, ZAM15,r<sup>-</sup>*

This *E. coli* TGI strain has the peculiarity of enabling recombinants to be recognized easily. The blue colour of the cells transfected with plasmids which did

not recombine with a fragment of LAV DNA is not modified. To the contrary cells transfected by a recombinant plasmid containing a LAV DNA fragment yield white colonies. The technique which was used is disclosed in Gene (1983), 26.

5 101.

This strain was transformed with the ligation mix using the Hanahan method (Hanahan D (1983) J. Mol. Biol. 166, 557). Cells were plated out on tryptone-agarose plate with IPTG and X-gal in soft agarose. White plaques were  
10 either picked and screened or screened directly in situ using nitrocellulose filters. Their DNAs were hybridized with nick-translated DNA inserts of pUC18 Hind III subclones of AJ19. this permitted the isolation of the plasmids or subclones of  $\lambda$  which are identified in the  
15 table hereafter. In relation to this table it should also be noted that the designation of each plasmid is followed by the deposition number of a cell culture of E. coli TGI containing the corresponding plasmid at the "Collection Nationale des Cultures de Micro-organismes" (C.N.C.M.) of  
20 the Pasteur Institute in Paris, France. A non-transformed TGI cell line was also deposited at the C.N.C.M. under Nr. I-364. All these deposits took place on November 15, 1984. The sizes of the corresponding inserts derived from the LAV genome have also been indicated.

25



TABLE

## Essential features of the recombinant plasmids

5	- pJ19 - 1 plasmid	(I-365)	0.5 kb
	Hind III - Sac I - Hind III		
10	- pJ19 - 17 plasmid	(I-367)	0.6 kb
	Hind III - Pst I - Hind III		
	- pJ19 - 6 plasmid	(I-366)	1.5 kb
15	Hind III (5')		
	Bam HI		
	Xho I		
	Kpn I		
	Bgl II		
20	Sac I (3')		
	Hind III		
	- pJ19-13 plasmid	(I-368)	6.7 kb
25	Hind III (5')		
	Bgl II		
	Kpn I		
	Kpn I		
	Eco RI		
30	Eco RI		
	Sal I		
	Kpn I		
	Bgl II		
	Bgl II		
35	Hind III (3')		

Positively hybridizing M13 phage plates were grown up for 5 hours and the single-stranded DNAs were extracted.

M13mp8 subclones of AJ19 DNAs were sequenced according to the dideoxy method and technology devised by Sanger et al (Sanger et al (1977), Proc. Natl. Acad. Sci. USA, 74, 5463 and M13 cloning and sequencing handbook, AMERSHAM (1983). the 17-mer oligonucleotide primer  $\alpha$ -<sup>35</sup>SdATP (400Ci/mmol, AMERSHAM), and 0.5X-5X buffer gradient gels (Biggen M.D. et al (1983, Proc. Natl. Acad. Sci. USA, 50, 3983) were used. Gels were read and put into the computer under the programs of Staden (Staden R. (1982), Nucl. Acids Res. 10, 4731). All the appropriate references and methods can be found in the AMERSHAM M13 cloning and sequencing handbook.

The complete sequence of AJ19 was deduced from the experiments as further disclosed hereafter.

Figs. 4-12 provide the DNA nucleotide sequence of the complete genome of LAV. The numbering of the nucleotides starts from a left most Hind III restriction site (5'AAG...) of the restriction map. The numbering occurs in tens whereby the last zero number of each of the numbers occurring on the drawings is located just below the nucleotide corresponding to the nucleotides designated. I.e. the nucleotide at position 10 is T, the nucleotide at position 20 is C, etc..

Above each of the lines of the successive nucleotide sequences there are provided three lines of single letters corresponding to the aminoacid sequence deduced from the DNA sequence (using the genetic code) for each at the three reading phases, whereby said single letters have the following meanings.

A : alanine  
R : arginine  
K : lysine  
H : histidine  
C : cysteine

M : méthionine  
 W : tryptophan  
 F : phenylalanine  
 Y : tyrosine  
 5 L : leucine  
 V : valine  
 I : isoleucine  
 G : glycine  
 T : thréonine  
 10 S : serine  
 E : glutamic acid  
 D : Aspartic acid  
 N : asparagine  
 Q : glutamine  
 15 P : proline.

The asterik signs "\*" correspond to stop codons (i.e. TAA, TAG and TGA).

Starting above the first line of the DNA nucleotidic sequence of fig. 4 the three reading phases are respectively marked "1", "2", "3", on the left handside of the drawing. The same relative presentation of the three theoretical reading phases is then used all over the successive lines of the LAV nucleotidic sequence.

Figs. 2 and 3 provide a diagrammatized representation of the lengths of the successive open reading frames corresponding to the successive reading phases (also referred to by numbers "1", "2" and "3" appearing in the left handside part of fig. 2). The relative positions of these open reading frames (ORF) with respect to the nucleotidic structure of the LAV genome is referred to by the scale of numbers representative of the respective positions of the corresponding nucleotides in the DNA sequence. The vertical bars correspond to the positions of the corresponding stop codons.

#### 35 1) The "gag gene" (or ORF-gag)

The "gag gene" codes for core proteins.

Particularly it appears that a genomic fragment (ORF-gag) thought to code for the core antigens including the p25, p18 and p13 proteins is located between nucleotidic position 236 (starting with 5' CTA GC6 GA6 3') and  
 5 nucleotidic position 1759 (ending by CTC6 TCA CAA 3'). The structure of the peptides or proteins encoded by parts of said ORF is deemed to be that corresponding to phase 2.

The methionine aminoacid "M" coded by the ATG at position 260-262 is the probable initiation methionine of  
 10 the gag protein precursor. The end of ORF-gag and accordingly of gag protein appears to be located at position 1759.

The beginning of p25 protein, thought to start by a P-I-V-Q-N-I-Q-G-Q-M-V-H .... aminoacid sequence is  
 15 thought to be coded for by the nucleotidic sequence CCTATA..., starting at position 658.

Hydrophilic peptides in the gag open reading frame are identified hereafter. They are defined starting from aminoacid 1 = Met (M) coded by the ATG starting from 260-2  
 20 in the LAV DNA sequence.

Those hydrophilic peptides are  
 12-32 aminoacids inclusive

	37-46	"	"
	49-79	"	"
25	88-153	"	"
	158-165	"	"
	178-188	"	"
	200-220	"	"
	226-234	"	"
30	239-264	"	"
	288-331	"	"
	352-381	"	"
	377-390	"	"
	399-432	"	"
35	437-484	"	"
	492-498	"	"

The invention also relates to any combination of these peptides.

2) The "pol gene" (or ORF-pol)

Figs. 4-12 also show that the DNA fragments  
5 extending from nucleotidic position 1555 (starting with  
5'TTT TTT ....3' to nucleotidic position 5086 is thought  
to correspond to the pol gene. The polypeptidic structure  
of the corresponding polypeptides is deemed to be that  
corresponding to phase 1. It stops at position 4563 (and  
10 by 5'G GAT GAG GAT 3').

These genes are thought to code for the virus  
polymerase or reverse transcriptase.

3) The envelope gene (or ORF-env)

The DNA sequence thought to code for envelope  
15 proteins is thought to extend from nucleotidic position  
5670 (starting with 5'AAA GAG GAG A....3') up to nucleo-  
tidic position 8132 (ending by ....A ACT AAA GAA 3').  
Polypeptidic structures of sequences of the envelope  
protein correspond to those read according to the "phase  
20 3" reading phase.

The start of env transcription is thought to be at  
the level of the ATG codon at positions 5691-5693.

Additional feature of the envelope protein coded  
by the env genes appear on figs. 13-18. These are to be  
25 considered as paired figs. 13 and 14 ; 15 and 16 ; 17 and  
18 respectively.

It is to be mentioned that because of format  
difficulties.

Fig. 14 overlaps to some extent with fig. 13.

30 Fig. 16 overlaps to some extent with fig. 15.

Fig. 18 overlaps to some extent with fig. 17.

Thus for instance figs. 13 and 14 must be con-  
sidered together. Particularly the sequence shown on the  
first line on the top of fig. 13 overlaps with the  
35 sequence shown on the first line on the top of fig. 14. In  
other words the starting of the reading of the successive

sequences of the env gene as represented in figs. 13-18 involves first reading the first line at the top of fig. 13 then proceeding further with the first line of fig. 14. One then returns to the beginning of the second line of fig. 13, then again further proceed with the reading of the second line of page 14, etc... The same observations then apply to the reading of the paired figs. 15 and 16, and paired figs. 17 and 18, respectively.

The locations of neutralizing epitopes are further apparent in figs. 13-18. reference is more particularly made to the boxed groups of three letters included in the aminoacid sequences of the envelope proteins (reading phase 3) which can be designated generally by the formula N-X-S or N-X-T, wherein X is any other possible aminoacid. Thus the initial protein product of the env gene in a glycoprotein of molecular weight in excess of 91,000. These groups are deemed to generally carry glycosylated groups. These N-X-S and N-X-T groups with attached glycosylated groups form together hydrophylic regions of the protein and are deemed to be located at the periphery of and to be exposed outwardly with respect to the normal conformation of the proteins. Consequently they are considered as being epitopes which can efficiently be brought into play in vaccine compositions.

The invention thus concerns with more particularly peptide sequences included in the env-proteins and excizable therefrom (or having the same aminoacid structure), having sizes not exceeding 200 aminoacids.

Preferred peptides of this invention (referred to hereafter as a, b, c, d, e, f) are deemed to correspond to those encoded by the nucleotide sequences which extend respectively between the following positions :

- a) from about 6095 to about 6200
- b) " " 6260 " " 6310
- 35 c) " " 6390 " " 6440
- d) " " 6485 " " 6620

e) " " 6860 " " 6930

f) " " 7535 " " 7630

Other hydrophilic peptides in the env open reading frame are identified hereafter. they are defined starting from

aminoacid 1 = lysine (K) coded by the AAA at position 5670-2 in the LAV DNA sequence.

These hydrophilic peptides are

8-23 aminoacids inclusive

10	63-78	"	"
	82-90	"	"
	97-123	"	"
	127-183	"	"
	197-201	"	"
15	239-294	"	"
	300-327	"	"
	334-381	"	"
	397-424	"	"
	488-500	"	"
20	510-523	"	"
	551-577	"	"
	594-603	"	"
	621-630	"	"
	657-679	"	"
25	719-758	"	"
	780-803	"	"

The invention also relates to any combination of these peptides.

#### 4) The other ORF

30 The invention further concerns DNA sequences which provide open reading frames defined as ORF-Q, ORF-R and as "1", "2", "3", "4", "5", the relative position of which appears more particularly in figs. 2 and 3.

These ORFs have the following locations :

35	ORF-Q	phase 1	start 4478	stop 5086
	ORF-R	" 2	" 8249	" 8896

ORF-1	"	1	"	5029	"	5316
ORF-2	"	2	"	5273	"	5515
ORF-3	"	1	"	5383	"	5616
ORF-4	"	2	"	5519	"	5773
5 ORF-5	"	1	"	7966	"	8279

The LTR (long terminal repeats) can be defined as lying between position 8560 and position 160 (end extending over position 9097/1). As a matter of fact the end of the genome is at 9097 and, because of the LTR structure of the retrovirus, links up with the beginning of the sequence :

Hind III  
CTCAATAAAGCTTGCCTTG

9097 1

The invention concerns more particularly all the DNA fragments which have been more specifically referred to hereabove and which correspond to open reading frames. It will be understood that the man skilled in the art will be able to obtain them all, for instance by cleaving an entire DNA corresponding to the complete genome of a LAV species, such as by cleavage by a partial or complete digestion thereof with a suitable restriction enzyme and by the subsequent recovery of the relevant fragments. The different DNAs disclosed in the earlier mentioned British Application can be resorted to also as a source of suitable fragments. The techniques disclosed hereabove for the isolation of the fragments which were then included in the plasmids referred to hereabove and which were then used for the DNA sequencing can be used.

Of course other methods can be used. Some of them have been exemplified in the earlier British Application. reference is for instance made to the following methods.

a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, calcium phosphate precipitation, polyethylene

glycol, protoplast-fusion, etc..

b) DNA fragments corresponding to genes can be cloned into expression vectors for *E. coli*, yeast- or mammalian cells and the resultant proteins purified.

5 c) The proviral DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinant producing antigenically competent fusion proteins can be identified by simply screening the recombinants with antibodies against LAV  
10 antigens.

The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to this invention, thus to recombinant DNAs containing such fragments, particularly any plasmids  
15 amplifiable in procaryotic or eucaryotic cells and carrying said fragments.

Using the cloned DNA fragments as a molecular hybridization probe - either by marking with radionucleotides or with fluorescent reagents - LAV virion RNA may be  
20 detected directly in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as Factor VIII concentrates) and vaccines, i.e. hepatitis B vaccine. It has already been shown that whole virus can be detected in culture supernatants of LAV producing cells. A  
25 suitable method for achieving that detection comprises immobilizing virus onto said a support e.g. nitrocellulose filters, etc., disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such an approach has already been  
30 developed for Hepatitis B virus in peripheral blood (according to SCOTTO J. et al. Hepatology (1983), 3, 379-384).

Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue  
35 of patients with LAV related symptoms, to see if the proviral DNA or RNA is present in host tissue and other

tissues.

A method which can be used for such screening comprise the following steps : extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV proviral DNA. Hybridization in situ can also be used.

Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionary related retrovirus exist. The methods referred to hereabove can be used, although hybridization and washings would be done under non stringent conditions.

The DNA according to the invention can be used also for achieving the expression of LAV viral antigens for diagnostic purposes.

The invention also relates to the polypeptides themselves which can be expressed by the different DNAs of the inventions, particularly by the ORFs or fragments thereof, in appropriate hosts, particularly procaryotic or eucaryotic hosts, after transformation thereof with a suitable vector previously modified by the corresponding DNAs.

These polypeptides can be used as diagnostic tools, particularly for the detection of antibodies in biological media, particularly in sera or tissues of persons afflicted with pre-AIDS or AIDS, or simply carrying antibodies in the absence of any apparent disorders. Conversely the different peptides according to this invention can be used themselves for the production of antibodies, preferably monoclonal antibodies specific of the different peptides respectively. For the production of hybridomas secreting said monoclonal antibodies conventional production and screening methods are used. These monoclonal antibodies, which themselves are part of

the invention then provide very useful tools for the identification and even determination of relative proportions of the different polypeptides or proteins in biological samples, particularly human samples containing  
5 LAV or related viruses.

Thus all of the above peptides can be used in diagnostics as sources of immunogens or antigens free of viral particles, produced using non-permissive systems, and thus of little or no biohazard risk.

10 The invention further relates to the hosts (procar-  
yotic or eucaryotic cells) which are transformed by the above mentioned recombinants and which are capable of expressing said DNA fragments.

Finally it also relates to vaccine compositions  
15 whose active principle is to be constituted by any of the expressed antigens, i.e. whole antigens, fusion polypep-  
tides or oligopeptides in association with a suitable pharmaceutical or physiologically acceptable carrier.


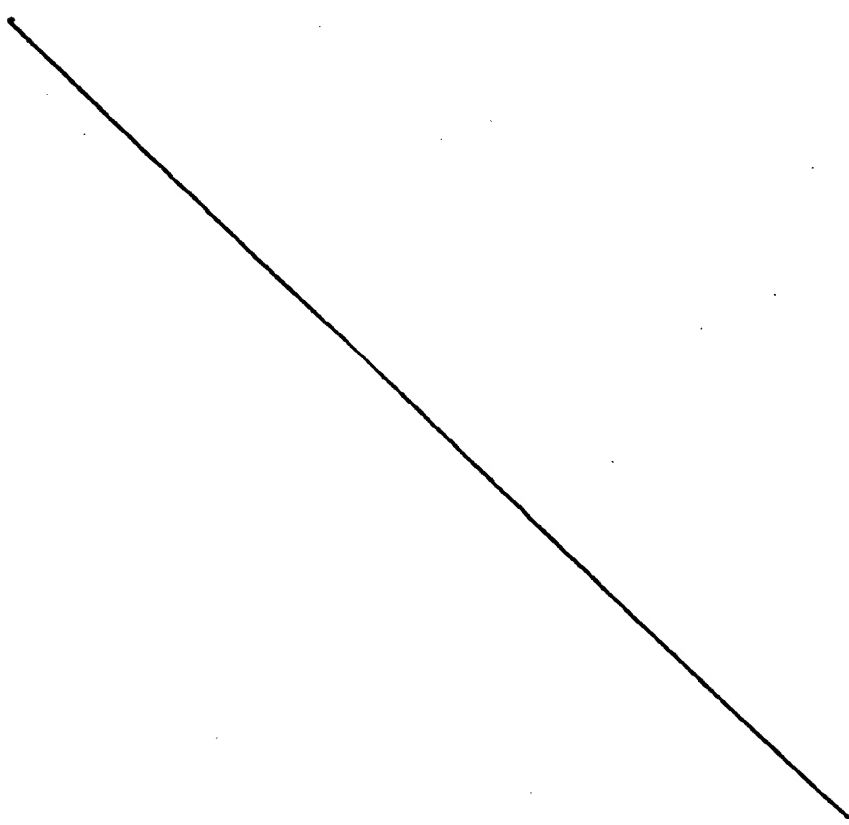
Preferably the active principles to be considered  
20 in that field consist of the peptides containing less than 250 aminoacid units, preferably less than 150 as deducible for the complete genomes of LAV, and even more preferably those peptides which contain one or more groups selected from N-X-S and N-X-T as defined above. Preferred peptides  
25 for use in the production of vaccinating principles are peptides (a) to (f) as defined above. By way of example having no limitative character, there may be mentioned that suitable dosages of the vaccine compositions are those which enable administration to the host,  
30 particularly human host ranging from 10 to 500 micrograms per kg, for instance 50 to 100 micrograms per kg.

For the purpose of clarity figs. 19 to 26 are added. reference may be made thereto in case of difficul-  
ties of reading blurred parts of figs. 4 to 12.

Needless to say that figs. 19-25 are merely a reiteration of the whole DNA sequence of the LAV genome.

Finally the invention also concerns vectors for the transformation of eucaryotic cells of human origin, particularly lymphocytes, the polymerases of which are capable of recognizing the LTRs of LAV. Particularly said vectors are characterized by the presence of a LAV LTR therein, said LTR being then active as a promoter enabling the efficient transcription and translation in a suitable host of the above defined, of a DNA insert coding for a determined protein placed under its controls.

Needless to say that the invention extends to all variants of genomes and corresponding DNA fragments (ORFs) having substantially equivalent properties, all of said genomes belonging to retroviruses which can be considered as equivalents of LAV.



CLAIMS

1. A DNA fragment of LAV extending from nucleotide position 236 to nucleotide position 1759.

2. A DNA fragment of LAV extending from nucleotide position 1555 to nucleotide position 5086.

3. A DNA fragment of LAV extending from nucleotide position 5670 to nucleotide position 8132.

4. A vector containing a DNA fragment according to any of claims 1 to 3.

5. Peptide corresponding to any of those encoded by the nucleotide sequences which extend respectively between the following positions :

a) from about 6095 to about 6200

b) " " 6260 " " 6310

c) " " 6390 " " 6440

d) " " 6485 " " 6620

e) " " 6860 " " 6930

f) " " 7535 " " 7630

6. Peptide characterized by a sequence of amino-acids deducible from LAV DNA the terminal aminoacids of which extend between the following positions with respect to the lysine (position 1) coded by the AAA at position 5670-5672 in the LAV DNA.

8-23 aminoacids inclusive

63-78 " "

82-90 " "

97-123 " "

127-183 " "

197-201 " "

239-294 " "

300-327 " "

334-381 " "

397-424 " "

466-500 " "

510-523 " "

551-577 " "

594-603 - -  
 621-630 - -  
 657-678 - -  
 719-758 - -  
 780-803 - -

or any combination of these peptides.

7. Peptide corresponding to the aminoacid sequences deducible from LAV DNA and the terminal aminoacids of which are positionned at the positions hereafter counted from the Met at position 1 coded by the ATG sequence at nucleotide positions 260-2 :

12-32 aminoacids inclusive

37-46 - -  
 49-79 - -  
 88-153 - -  
 158-165 - -  
 178-188 - -  
 200-220 - -  
 226-234 - -  
 239-264 - -  
 288-331 - -  
 352-361 - -  
 377-390 - -  
 399-432 - -  
 437-484 - -  
 492-498 - -

and combination of said peptides.

8. Diagnostic means containing any of the DNA fragments of any of claims 1 to 3.

9. Diagnostic means containing any of the peptides of any of claims 4 to 6.

10. Vaccine compositions containing any of the peptides according to any of claims 4 to 6 in association with a pharmaceutical vehicle.

End of transmission

[illegible]

DFA 2/26



Fig. 2

16 NOV. 84- 29099

DFA 3/26

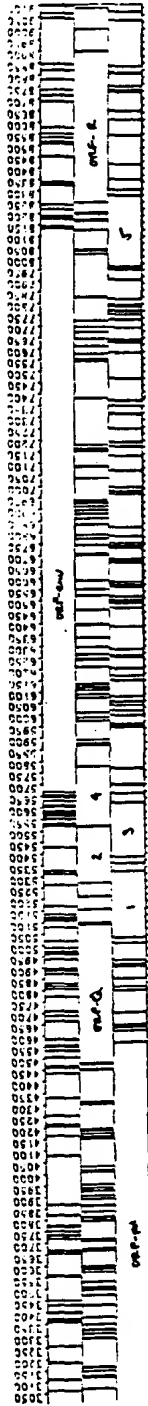


Fig. 3

DFA 4/26

Fig. 4

Fig. 5

Full length

V V K L K P G M D G P K V V Q A P L T F F K I K A L V I C T I V E K E G R I

3

V V K L K P G M D G P K V V Q A P L T F F K I K A L V I C T I V E K E G R I  
2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160

2

S K I G P E H P V V T P V F A I V V D S T K P K L V D F V E L V N P T O D F  
2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280

1

H F V Q L G I P H P A G L K K K K S V T V L G V G D A T F S V P L U E D F R K V  
2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

1

L A F T L P S I H H F T P G I R V Q V V V L P G G A V S P A I F U S S M T K I  
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

1

L F P F P Q Q N P O I V I V U T M D Q L V V C S D L F I G C N Q T K I E F L N O  
2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

1

N L L G M S L T T P O R V H L Q I P P F L A P G V E L V P O R A T V Q P I V L P  
2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760

1

F Q D S V T V U I G K L V G K L V M A S Q I V P G I K V Q Q V L V L R G I K  
2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880

1

A L T F V I P L T E I A F L L G V Q I I L Q E P V M S V V F P S Y D L I A  
2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000

Fig 6

[illegible]

Fig 7



D F A

9/26

[illegible]

15





12

S N H K O U V S Y O C C L C L A R S T O G C G C G T S M T S G T F K Y M D L  
 A I T S M I A T : A A C C A U L F A O E E F E V G F P V T P C V P L R P M L T  
 O S O V A I U J P L P L V P C O K M K R R M V F C S M L R Y L O O O L T  
 G C C A T C A G T A G C A C C A G C T A C C A A T G C T T G C T G C C T A G C A C C A G C A G G T G C G T T C C A G C A C C A G C T G C T T T A G C A C C A G T A  
 4430 4435 4440 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920 4930 4940 4950 4960 4970 4980 4990 5000

U  
G  
C  
S  
O  
P  
T  
E  
N  
M  
G  
Y  
G  
A  
N  
J  
I  
R  
V  
O  
S  
V  
U  
H  
F  
J

2 K A A V D L S H F L M E K C G L I M S O P Q R O D I L D L M I Y T O G V F  
R O L I L A T F + K K R C D M K G F T P W E D K I L L I C G T C T A C T  
CAAGCGAGCTGTACATTCACCATTTTAAAGAAAGGGGAGCTGCAAGGGCTTATCTCCCAAGACAGATATCTGTCTGCTATCCACACCAAGCTCT  
#530 #550 #570 #590 #610 #630 #650

POLAE L M T P R A C Q I S T D L W H V L O A S T S \* A R \* G R R C Z \* R P F  
P P D M O E N Y T P C P C V R V P L T F G W C Y K L V P L S O L R \* E R P I M E R T  
L I G K T Y T H J C U C A T C H \* P L D C A T S \* Y O L S \* O L R \* E R P I M E R T  
C C C G A T T G C G A G A C T A C A C A C G G C C G C G C G T C A C A T A T C C C T C A C G T A C T A C G A C T C C G A G A C C C A A A G A G A G A A  
8690 8680 8670 8660 8650 8640 8630 8620 8610 8600 8590 8580 8570 8560 8550 8540 8530 8520 8510 8500

W J L V T A C E P A M C O P E R S V R V F V O P S I S S R C P R A A S P  
 T S L L M P V S L H C D U P F R E V L C G B F D S R L A F M H V A R E L M P R S  
 P A C Y T L A C T F L R F C O S C G L T A A H F I T V P R E S C I R S  
 C A C C G C T T T A C C C G C C G C C A T C G A C A G A G C T T T A C A C C C C C T A C C C C C T A C C C C C C A G A G C T C A C C G G A  
 8770 8768 8790 8770 8820 8810 8900 8930 8940 8950 8960 8970 8980

[illegible][illegible]

19092

fig 12

16 NOV 84 29099 131  
 D F A  
 CAGAGGAGAGCAAGAAATCGAGGAGATAGATCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCAT  
 5290 5300 5310 5320 5330 5340 5350

16 NOV 84-729099

S O P K T A C T T C Y C K K C C F H C  
V S L K L L V P L A I V K S V A F I A  
CAGGAAGTCAGCCTAAACTGCTTGTACCCTTGTCTATTGTAAAAAGTGTGCTTTTCATTG  
5350 5360 5370 5380 5390 5400

A T K T S S P O S D S S S F S I K A V S  
O R R R P P Q G S G T H Q V S L S K O \* V  
S D E D L L K A V R L I K F L Y O S S K \*  
AGCGACGAAGACCTCCTCAAGGCAGTCAGACTCATCAAGTTTCTCTATCAAAGCAGTAAGT  
5470 5480 5490 5500 5510 5520

S N S C V V H S N H R I \* E N I K T K K  
I A I V V W S I V I I E Y R K I L R O R K  
\* O \* L C G P \* \* S \* N I G K Y \* D K E K  
TAGCAATAGTTGTGTGGTCCATAGTAATCATAGAATATAGGAAAATATTAAGACAAAGAAA  
5590 5600 5610 5620 5630 5640

R R N I S T C G D G G G N G A P C S L G  
G E I S A L V E M G V E M G H H A P W D  
K E K Y Q H L W R W G K W G T M L L G I  
AAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGA  
5710 5720 5730 5740 5750 5760

C G F K Q P P L Y F V H O M L K H M I O  
V E G S N H H S I L C I R C \* S I \* Y R  
V W K E A T T T L F C A S D A K A Y D T E  
TGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAG  
5830 5840 5850 5860 5870 5880

\* Y W \* M \* O K I L T C G K M T W \* N R  
S I G K C D R K F \* H V E K \* H G R T D  
V V L V N V T E N F N M W K N D H V E O M  
TAGTATTGGTAAATGTGACAGAAAATTTAACATGTGAAAAATGACATGGTAGAACAGA  
5950 5960 5970 5980 5990 6000

H S V L V \* S A L T W G \* L L I P I V V  
T L C \* F K V H \* F G E C Y \* Y O \* \* \*  
P L C V S L K C T D L G N A T N T N S S N  
CACTCTGTGTAGTTTAAAGTGCACTGATTTGGGCAATGCTACTAATACCAATAGTAGTA  
6070 6090 6090 6100 6110 6120

S I S A O A \* E V R C P K N M H F F I N  
O Y O H K H K R \* G A E R I C I F L \* T  
P N I S T S I R G K V Q K E Y A F F Y K L  
TCAATATCAGCACAAGCATAAGAGGTAAGGTCCAGAAAGAAATATGCATTTTTTATAAAC  
6190 6200 6210 6220 6230 6240

O S L H R P V Q R Y P L S O F P Y I I V  
S H Y T G L S K G I L \* A N S H T L L C  
S V I T O A C P K V S F E P I P I H Y C A  
CAGTCATTACACAGGCTGTCCAAAGGTATCCTTTGAGCCAATTCACATACATTATTGTC  
6310 6320 6330 6340 6350 6360

V O M S A O Y N V H \* F L G O \* Y O L N

16 NOV. 84 29039

DEF 18

P G A F C D S K \* | \* \* V D W H R T M Y K C Q  
P A G F A I L K C H [N K T] F [N G T] G P C T [Y V S]  
CCCCGGCTGTTTGGCATTCTAAAATGTAATAAAGACGTTCAATGGAACAGGACCATGTACAAATGTACG  
6370 6390 6390 6400 6410 6420 6430

C C \* H A V \* Q K K R \* \* L D L P I S O T M L K I  
A V E W O S S R R R G S N \* I C O F H R O C \* N  
L L [N G S] L A E E E V V I R S A [N F T] D N A K T  
TGCTGTTGAATGGCACTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCTAAAACC  
6490 6500 6510 6520 6530 6540 6550

P T T I Q E K V S V S R G D O G E H L L O \* E K \*  
Q Q Q Y K K K Y P Y P E G T R E S I C Y N P K N  
N [N N T] R K S I R I O R G P G R A F V T I G K I  
CCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTTACAAATAGGAAAAATA  
6610 6620 6630 6640 6650 6660 6670

M P L \* N R \* L A N \* E N N L E I I K O \* S L S N  
C H F K T D S \* Q I K R T I W K \* \* N N N L \* A  
[A T] L K Q I A S K L R E O F G N [N K T] I I F K Q  
ATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAA  
6730 6740 6750 6760 6770 6780 6790

I G N F S T V I O H N C L I V L G L I V L G V L K  
K G I F L L \* F N T T V \* \* Y L V \* \* Y L E Y \*  
G E F F Y C [N S T] Q L F [N S T] W F [N S T] W S T E  
CAGGGGAATTTTCTACTGTATTCAACACAACCTGTTAATAGTACTTGGTTAATAGTACTTGGAGTACTGAA  
6850 6860 6870 6880 6890 6900 6910

E \* N N L \* T C G R K \* E K Q C M P L P S A D K L  
N K T I Y K H V A G S R K S N V C P S H O R T H  
I K O F I N M H O E V G K A M Y A P P I S G Q I  
GAATAAAACAATTTATAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATT  
6970 6980 6990 7000 7010 7020 7030

V I T T H G P R S S D L E E E I \* G T I G E V N Y  
\* \* O O W V R D L O T W R R R Y E G O L E K \* I  
N N N [N G S] E I F R P G G G G D M R O N W R S E L  
GTAATAACAACAATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTA  
7090 7100 7110 7120 7130 7140 7150

P R Q R E E W C R E K K E O W E \* E L C S L G S W  
O G K E K S G A E R K K S S G N R S F V P W V L  
K A K R R V V O R E K R A V G I G A L F L G F L  
CCAAGGCAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGT  
7210 7220 7230 7240 7250 7260 7270

Y R P D N Y C L V \* C S S R T I C \* G L L R R N S  
T G O T I I V W Y S A A E Q F A E G Y \* G A T  
O A R O L L S G I V O O Q O N N L R A I E A O Q  
TACAGGCCAGACAATTATTGCTGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGCGGCAACAGC  
7330 7340 7350 7360 7370 7380 7390

E S A L W K D T \* R I N S S W G F G V A L E N S F

DFA

N T M Y K C Q H S T M Y T M N \* A S S I V S  
 T G P C T N V S T V O C T H G I R P V V S T U L  
 AACAGGACCATGTACAAATGTCAGCACAGTACAAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAAC  
 6420 6430 6440 6450 6460 6470 6480

P I S O T M L K P \* \* Y S \* T V L \* K L I V U D  
 O F H R O C \* N H N S T A E P I C R N \* L Y K T  
 N E T D N A K T I I V O L N O S V E I N C T R P  
 CAATTTACAGACAATGCTAAAACATAATAGTACAGCTGAACCAATCTGTAGAATTAATTGTACAAGAC  
 6540 6550 6560 6570 6580 6590 6600

F H L L O \* E K \* E I \* D K H I V T L V F O N G  
 S I C Y V P K N R K Y E T S T L \* H \* \* S K M E  
 A F V T I G K I G N M R O A H C N I S R A K W N  
 AGCATTGTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGA  
 6660 6670 6680 6690 6700 6710 6720

I I K O \* S L S N P O E G T O K L \* R T V L I V  
 \* \* N N N L \* A I L R R G P R N C N A O F \* L W  
 N K T I I F K O S S S G G O P E I V T H S F N C G  
 TAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTG  
 6780 6790 6800 6810 6820 6830 6840

G L I V L G V L K G Q I T L K E V T O S H S H A  
 V \* \* Y L E Y \* R V K \* H \* H N H T P M R  
 F N S T W S T E G S N N T E G S O T I T L P C R  
 GTTAATAGTACTTGGAGTACTGAAGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCA  
 6900 6910 6920 6930 6940 6950 6960

N P L P S A D K L D V H O I L O G C Y \* O E M V  
 C P S H O R T N \* M F I K Y Y R A A I N K R W H  
 A P P I S G O I R C S S N I T G L L L T R D G G  
 TGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAAATATTACAGGCTGCTATTAACAAGAGATGGTG  
 7020 7030 7040 7050 7060 7070 7080

\* G T I G E V N Y I N I K \* \* K L N H \* E \* H P  
 E G O L E K \* I I \* I \* S S K N \* T I R S S T H  
 R D N W R S E L Y K Y K V V K I E P L G V A P T  
 CAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATAGGAGTAGCACCCA  
 7140 7150 7160 7170 7180 7190 7200

\* E L C S L G S W E O O E A L W A H G O \* R \* R  
 R S F V P W V L G S S R K H Y G R T V N O A O G  
 G A L F L G F L G A A G S T M G A R S M T L T V  
 AGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGGTCAATGACGCTGACGG  
 7260 7270 7280 7290 7300 7310 7320

C \* G L L R R N S I C C N S O S G A S S S S R O  
 A E G Y \* G A T A S V A T H S L G H O A A P G K  
 L R A I E A O O H L L O L T V W G I K O L O A R  
 CTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAA  
 7380 7390 7400 7410 7420 7430 7440

G V A L E N S F A P L L C L G \* L V G V I N L

16 NOV. 84-29899

N P C C G K I P K G S T A P G D L G L L A K T H  
I L A V E R Y L K D O U L L G I W C C S G K L I  
GAATCCTGGCTGTGGAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAGAACTCATI  
7450 7460 7470 7480 7490 7500 7510

W N R F G I T \* P G W S G T E K L T I T Q A \* Y I  
G T D L E \* H D L D G V G O R N \* O L H K L N T  
E O I W N N T W M E W D R E I N N Y T S L I H  
TGAACAGATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACAI  
7570 7580 7590 7600 7610 7620 7630

N Y W N \* I N G O V C S I G L T \* O I G C G I \* K  
I I G I R \* M G K F V E L V \* H N K L A V V Y K  
L L E L D K W A S L W N W F N I T N W L W Y I K  
AATTATTGGAATTAGATAAATCGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTGGCTGTGTTATATAAAA  
7690 7700 7710 7720 7730 7740 7750

L L Y F L \* \* I E L G R D I H H Y R F R P T S Q P  
C C T F Y S E \* S \* A G I F T I I V S D P P P H  
A V L S I V / N R V R O G Y S P L S F O T H L P T  
TTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTTATCGTTTCAGACCCACCTCCCAACC  
7810 7820 7830 7840 7850 7860 7870

R E T E T D P F D \* \* T D P \* H L S G T I C G A L  
E R U R O I H S I S E R I L S T Y L G R S A E P  
R D R D R S I R L V N G S L A L I W D D L R S L  
AGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTCGGACGATCTCGGAGCCTT  
7930 7940 7950 7960 7970 7980 7990

T R I V E L L G K R G W E A L K Y W W N L L O Y W  
R G L W N F W D A G G G K P S N I G G I S Y S I  
E D C G T S G T G T Q G V G S P O I L V E S P T V L  
ACGAGGATTGTGGAACCTTCTGGGACGCGGGTGGGAAGCCCTCAAATATTGGTGAATCTCTACAGATTG  
8050 8060 8070 8080 8090 8100 8110

A I A V A E G T D R V I E V V O G A C R A I R H I  
P \* J \* L R G O I G L \* K \* Y K E L V E L F A T  
H S S S \* G D R \* G Y R S S T R S L \* S Y S P H  
GCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTAGTACAAGGACCTGTAGAGCTATTGCGCACAT  
8170 8180 8190 8200 8210 8220 8230

G W O V Y K K \* C G W M A Y C K G K N E T S \* A S  
G G K \* S K S S V V G W P T V R E R M R R A E P  
Y A S G O K V V W L D G L L \* G K E \* D E L S O  
GGGTGGCAAGTGGTCAAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGGAAGAATGAGACGAGCTGAGCCAG  
8290 8300 8310 8320 8330 8340 8350

S N H K \* O Y S S Y O C C L C L A R S T R G G G C  
A I T S S N T A A T N A A C A W L F A O E E E E E  
O S O V A I O O L P M L L Y P G \* K H K R R R R  
ACCAATCAAGTAGCAATACAGCAGCTACCAATGCTGCTTGGCTGGCTGCAAGGACAGAGGAGGAGGAGG  
8410 8420 8430 8440 8450 8460 8470

U G S C R S \* P L F K R K C G T C  
8/15/15

10 NOV 04 23055  
13/26

13/18  
DFA

W K T H L H H C C A L E C \* L E \* \* I S  
K L I C T T A V P W W A S W S W A S L  
CTGGAAACTCATTTGCACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTC  
7510 7520 7530 7540 7550 7560

O A \* Y I P \* L K N R K T S K K R M N K  
K L N T F L N \* R I A K P A R K E \* T R  
S L I H S L I E E S O N O E K N E Q E  
CAAGCTTAATACATTCCTTAATTGAAGAATCGAAAACCAGCAAGAAAAGAATGAACAAG  
7630 7640 7650 7660 7670 7680

C G I \* K Y S \* \* \* \* E A W \* V \* E \* F  
V V Y K N I H N D S R R L G R F K N S F  
W Y I K I F I M I V G G L V G L / R / I V F  
TGTGGTATATAAAATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTT  
7750 7760 7770 7780 7790 7800

P T S O P R G D P T G P K E \* K K K V E  
P P P N P E G T R O A R R N R R R R W R  
T H L P T P R G P D R P E G I E E E G G E  
CCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAG  
7870 7880 7890 7900 7910 7920

I C G A L C L F S Y H R L R D L L L I V  
S A E P C A S S A T T A \* E T Y S \* L \*  
L R S L V P L O L P P L E R L T L D C N  
ATCTGCGGAGCCTTGTGCCTCTTCAGCTACCACCCCTTGAGAGACTTACTCTTGATTGTA  
7990 8000 8010 8020 8030 8040

L L O Y W S O E L K N S A V S L L N A T  
S Y S I G V R N \* R I V L L A C S M P O  
P T V L E S G T K E \* C C \* L A O C H S  
TCCTACAGTATTGGAGTCAGGAACATAAGAATAGTGCTGTTAGCTTGCTCAATGCCACA  
8110 8120 8130 8140 8150 8160

A I R H I P R R I R O G L E R I L L \* D  
L F A T Y L E E \* D R A W K G F C Y K M  
Y S P H T \* K N K T G L G K D F A I R W  
CTATTGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGTATAAGAT  
8230 8240 8250 8260 8270 8280

T S \* A S S R W G G S S I S R P G K T W  
R A E P A A D G V G A A S R D L E K H G  
E L S O O O \* G W E O H L E T W K N M E  
CGAGCTGAGCCAGCAGATGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATCG  
8350 8360 8370 8380 8390 8400

R G G G G G F S S H T S G T F K T N D L  
E E E E V G F P V T P C V P L R P M T Y  
R R P R W Y F S S H L R Y L \* D O \* L T  
JAGGAGGAGGAGGGGGTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTA  
8470 8480 8490 8500 8510 8520

Using track  
S L P T K T Y P \* S V O L P M T R L L  
15/15 B/L

16 NOV. 84- 29099

19/28

D F A

Fig 19

10	20	30	40	50	60
AAGCTTGCCT	TGAGTGCTTC	AAGTAGTGTG	TGCCCCGTCTG	TTGTGTGACT	CTGGTAACATA
70	80	90	100	110	120
GAGATCCCTC	AGACCCTTTT	AGTCAGTGTG	GAAAATCTCT	AGCAGTGCCG	CCCGAACAGG
130	140	150	160	170	180
GACTTGAAAG	CGAAAGGGAA	ACCAGAGGAG	CTCTCTCCAC	GCAGGACTCG	GCTTGCTGAA
190	200	210	220	230	240
GCGCGCACGG	CAAGAGGCGA	GGGGAGGCCA	CTGGTGACTA	CGCCAAAAAT	TTTGACTAGC
250	260	270	280	290	300
GGAGGCTAGA	AGGAGAGAGA	TGGGTCCGAG	AGCGTCAGTA	TTAAGCGGGG	GAGAATTAGA
310	320	330	340	350	360
TCGATGGGAA	AAAATTCGGT	TAAGGCCAGG	GGGAAAGAAA	AAATATAAAT	TAAACATAT
370	380	390	400	410	420
AGTATGGGCA	AGCAGGGAGC	TAGAACGATT	CGCTGTTAAT	CCTGGCCTGT	TAGAAACATC
430	440	450	460	470	480
AGAAGGCTGT	AGACAAATAC	TGGGACAGCT	ACAACCATCC	CTTCAGACAG	GATCAGAAGA
490	500	510	520	530	540
ACTTAGATCA	TTATATAATA	CAGTAGCAAC	CCTCTATTCT	GTGCATCAAA	GGATAGAGAT
550	560	570	580	590	600
AAAAGACACC	AAGGAAGCTT	TAGACAAGAT	AGAGGAAGAG	CAAAACAAAA	GTAAGAAAAA
610	620	630	640	650	660
AGCACAGCAA	GCAGCAGCTG	ACACAGGACA	CAGCAGCCAG	GTCAGCCAAA	ATTACCCTAT
670	680	690	700	710	720
AGTGCAGAAC	ATCCAGGGGC	AAATGGTACA	TCAGGCCATA	TCACCTAGAA	CTTTAAATGC
730	740	750	760	770	780
ATGGGTAAAA	GTAGTAGAAG	AGAAGGCTTT	CAGCCCAGAA	GTGATACCCA	TGTTTTTCAGC
790	800	810	820	830	840
ATTATCAGAA	GGAGCCACCC	CACAAGATTT	AAACACCATG	CTAAACACAG	TGGGGGGACA
850	860	870	880	890	900
TCAAGCAGCC	ATGCAAAATGT	TAAAAGAGAC	CATCAATGAG	GAAGCTGCAG	AATGGGATAG
910	920	930	940	950	960
AGTGCATCCA	GTGCATGCAG	GGCCTATTGC	ACCAGGCCAG	ATGACAGAAC	CAAGGGGAAG
970	980	990	1000	1010	1020
TGACATAGCA	GGAAGTACTA	GTACCCTTCA	GGAACAAATA	GGATGGATGA	CAATAATCC
1030	1040	1050	1060	1070	1080
ACCTATCCCA	GTAGGAGAAA	TTTATAAAAG	ATGGATAATC	CTGGGATTAA	ATAAAATAGT
1090	1100	1110	1120	1130	1140

DEB/26

Fig 20

AAGAATGTAT AGCCCTACCA GCATTCTGGA CATAAGACAA GGACCAAAAG AACCCCTTATG  
 1150 1160 1170 1180 1190 1200  
 AGACTATGTA GACCGGTTCT ATAAAACTCT AAGAGCCGAG CAAGCTTCAC AGGAGGTAAA  
 1210 1220 1230 1240 1250 1260  
 AAATTGGATG ACAGAAACCT TGTTCGTCCA AAATCCGAAC CCAGATTGTA AGACTATTTT  
 1270 1280 1290 1300 1310 1320  
 AAAAGCATTG GGACCAGCAG CTACACTAGA AGAAATGATG ACAGCATGTC AGGGAGTGGG  
 1330 1340 1350 1360 1370 1380  
 AGGACCCGGC CATAAGGCAA GAGTTTTGGC TGAAGCAATG AGCCAAGTAA CAAATTCAGC  
 1390 1400 1410 1420 1430 1440  
 TACCATAATG ATGCAAAGAG GCAATTTTAG GAACCAAGA AAGATTGTGA AGTGTTCAT  
 1450 1460 1470 1480 1490 1500  
 TTGTGGCAAA GAAGGGCACA TAGCCAGAAA TTGCAGGGCC CCTAGGAAAA AGGGCTGTTG  
 1510 1520 1530 1540 1550 1560  
 GAAATGTGGA AAGGAAGGAC ACCAAATGAA AGATTGTACT GAGAGACAGG CTAATTTTTT  
 1570 1580 1590 1600 1610 1620  
 AGGGAAGATC TGGCCTTCCT ACAAGGGAAG GCCAGGGAAT TTTCTTCAGA GCAGACCAGA  
 1630 1640 1650 1660 1670 1680  
 GCCAACAGCC CCACCAGAAG AGAGCTTCAG GTCTGGGGTA GAGACAACAA CTCCTCTCA  
 1690 1700 1710 1720 1730 1740  
 GAAGCAGGAG CCGATAGACA AGGAAGTGA TCCTTTAACT TCCCTCAGAT CACTCTTTGG  
 1750 1760 1770 1780 1790 1800  
 CAACGACCCC TCGTCACAAT AAAGATAGGG GGGCAACTAA AGGAAGCTCT ATTAGATACA  
 1810 1820 1830 1840 1850 1860  
 GGAGCAGATG ATACAGTATT AGAAGAAATG AGTTTGCCAG GAAGATGGAA ACCAAAAATG  
 1870 1880 1890 1900 1910 1920  
 ATAGGGGGAA TTGGAGGTTT TATCAAAGTA AGACAGTATG ATCAGATACT CATAGAAATC  
 1930 1940 1950 1960 1970 1980  
 TGTGGACATA AAGCTATAGG TACAGTATTA GTAGGACCTA CACCTGTCAA CATAATTGGA  
 1990 2000 2010 2020 2030 2040  
 AGAAATCTGT TGAAGTATG TGGTTGCACT TTAATTTTTC CCATTAGTCC TATTGAAACT  
 2050 2060 2070 2080 2090 2100  
 GTACCAGTAA AATTAAAGCC AGGAATGGAT GGCCCAAAAG TTAACAATG GCCATTGACA  
 2110 2120 2130 2140 2150 2160  
 GAAGAAAAAA TAAAAGCATT AGTAGAAAT TGTACAGAAA TGGAAAAGGA AGGGAAAATT  
 2170 2180 2190 2200 2210 2220  
 TCAAAAATTG GGCTGAAAA TCCATACAAT ACTCCAGTAT TTGCCATAAA GAAAAAAGAC  
 2230 2240 2250 2260 2270 2280  
 AGTACTAAAT GGAGAAAATT AGTAGATTTT AGAGAACTTA ATAAGAGAAC TCAAGACTTC  
 2290 2300 2310 2320 2330 2340  
 TGGGAAGTTC AATTAGGAAT ACCACATCCC GCAGGGTTAA AAAAGAAAAA ATCAGTAACA

10-84-29099

D F A

GTACTGGATG TGGGIGATGC ATATTTTICA GTTCCCTTAG ATGAAGACTT CAGGAAGIAT

2410 2420 2430 2440 2450 2460  
ACTGCATTTA CCATACCTAG TATAACAAT GAGACACCAG GGATTAGATA TCAGTACAAT

2470 2480 2490 2500 2510 2520  
GTGCTTCCAC AGGGATGGAA AGGATCACCA GCAATATTCC AAAGTAGCAT GACAAAAATC

2530 2540 2550 2560 2570 2580  
TTAGAGCCTT TTAGAAAAACA AAATCCAGAC ATAGTTATCT ATCAATACAT GGATGATTGG

2590 2600 2610 2620 2630 2640  
TATGTAGGAT CTGACTTAGA AATAGGGCAG CATAGAACAA AAATACAGGA GCTGAGACAA

2650 2660 2670 2680 2690 2700  
CATCTGTTGA GGTGGGGACT TACCACACCA GACAAAAAAC ATCAGAAAGA ACCTCCATTC

2710 2720 2730 2740 2750 2760  
CTTTGGATGG GTTATGAACT CCATCCTGAT AAATGGACAG TACAGCCTAT AGTGGCTGCCA

2770 2780 2790 2800 2810 2820  
GAAAAAGACA GCTGGACTGT CAATGACATA CAGAAGTTAG TGGGAAATTT CAATTGGGCA

2830 2840 2850 2860 2870 2880  
AGTCAGATTT ACCCAGGGAT TAAAGTAAGG CAATTATGTA AACTCCTTAG AGGAACCAAA

2890 2900 2910 2920 2930 2940  
GCACTAACAG AAGTAATACC ACTAACAGAA GAAGCAGAGC TAGAACTGGC AGAAAACAGA

2950 2960 2970 2980 2990 3000  
GAGATTCTAA AAGAACCAGT ACATGGAGTG TATTATGACC CATCAAAAGA CTTAATAGCA

3010 3020 3030 3040 3050 3060  
GAAATACAGA AGCAGGGGCA AGGCCAATGG ACATATCAAA TTTATCAAGA GCCATTTAAA

3070 3080 3090 3100 3110 3120  
AATCTGAAAA CAGGAAAAATA TGCAAGAACG AGGGGTGCCC AACTAATGA TGTAACAACA

3130 3140 3150 3160 3170 3180  
TTAACAGAGG CAGTGCAAAA AATAACCACA GAAAGCATAG TAATATGGCC AAAGACTCCT

3190 3200 3210 3220 3230 3240  
AAATTTAAAC TACCCATACA AAAGGAAACA TGGCAACAT GGTGGACAGA GTATTGGCAA

3250 3260 3270 3280 3290 3300  
GCCACCTGGA TTCCTGAGTG GGAGTTTGTC AATACCCCTC CTTAGTGAA ATTATGCTAC

3310 3320 3330 3340 3350 3360  
CAGTTAGACA AAGAACCCAT AGTAGGAGCA GAAACGTTCT ATGTAGATGG GGCAGCTAGC

3370 3380 3390 3400 3410 3420  
AGGGGAGACTA AATTAGGAAA AGCAGGATAT GTTACTAATA GAGGAAGACA AAAAGTTGTC

3430 3440 3450 3460 3470 3480  
ACCCTAACTG ACACAACAAA TCAGAAGACT GAGTTACAAG CAATTATCT AGCTTTGCAG

3490 3500 3510 3520 3530 3540  
GATTCGGGAT TAGAAGTAAA TATAGTAACA GACTCACAAT ATGCATTAGG AATCATTCAA

3550 3560 3570 3580 3590 3600  
GCACAACCAG ATAAAAGTGA ATCAGAGTTA GTCAATCAAA TAATAGAGCA GTTAATAAAA

3610 3620 3630 3640 3650 3660

221  
Fig 92

16 NOV 84 - 29099

3670	3680	D F A 3690	3710	3720
GTAGATAAAT	TAGTCAGTGC	TGGAATCAGG	AAAGTACTAT	TTTATAGTGG
3730	3740	3750	3760	3770
GCCCAAGATG	AACATGAGAA	ATATCACAGT	AATTGGAGAG	CAATGGCTAG
3790	3800	3810	3820	3830
CTGCCACCTG	TAGTAGCAAA	AGAAATAGTA	GCCAGCTGTC	ATAAATGTCA
3850	3860	3870	3880	3890
GAAGCCATGC	ATGGACAAGT	AGACTGTAGT	CCAGGAATAT	GGCAACTACA
3910	3920	3930	3940	3950
TTAGAAGGAA	AAGTTATCCT	GGTAGCAGTT	CATGTAGCCA	GTGGATATAT
3970	3980	3990	4000	4010
GTTATTCCAG	CAGAAACAGG	GCAGGAAACA	GCATACCTTC	TTTTAAATTT
4030	4040	4050	4060	4070
TGGCCAGTAA	AAACAATACA	TACAGACAAT	GCCAGCAATT	TCACCAGTAC
4090	4100	4110	4120	4130
GCCGCCTGTT	GGTGGGCGGG	AATCAAGCAG	GAATTTGGAA	TTCCCTACAA
4150	4160	4170	4180	4190
CAAGGAGTAG	TAGAATCTAT	GAATAAGAA	TTAAAGAAAA	TTATAGGCCA
4210	4220	4230	4240	4250
CAGGCTGAAC	ATCTTAAGAC	AGCAGTACAA	ATGGCAGTAT	TCATCCACAA
4270	4280	4290	4300	4310
AAAGGGGGGA	TTGGGGGGTA	CAGTGCAGGG	GAAAGAATAG	TAGACATAAT
4330	4340	4350	4360	4370
ATACAAACTA	AAGAATTACA	AAAACAAATT	ACAAAAATTC	AAAATTTTCG
4390	4400	4410	4420	4430
AGGGACAGCA	GAGATCCACT	TTGGAAAGGA	CCAGCAAAGC	TCCTCTGGAA
4450	4460	4470	4480	4490
GCAGTAGTAA	TACAAGATAA	TAGTGACATA	AAAGTAGTGC	CAAGAAGAAA
4510	4520	4530	4540	4550
ATTAGGGATT	ATGGAAAACA	GATGGCAGGT	GATGATTGTC	TGGCAAGTAG
4570	4580	4590	4600	4610
GATTAGAAC	TGGAAAAGTT	TAGTAAAACA	CCATATGTAT	GTTTCAGGGA
4630	4640	4650	4660	4670
ATGGTTTTAT	AGACATCACT	ATGAAAGCCC	TCATCCAAGA	ATAAGTTCAG
4690	4700	4710	4720	4730
CCCACTAGGG	GATGCTAGAT	TGGTAATAAC	AACATATTGG	GGTCTGCATA
4750	4760	4770	4780	4790
AGACTGGCAT	CTGGGTCAGG	GAGTCTCCAT	AGAATGGAGG	AAAAAGAGAT
4810	4820	4830	4840	4850
AGTAGACCCT	CAACTAGCAG	ACCAACTAAT	TCATCTGTAT	TACTTTGACT
4870	4880	4890	4900	4910

16 NOV 84 - 29099  
 4930 4940 D F A4950 4960 4970 4980  
 AGGACATAAC AAGGTAGGAT CTCTACAATA CTTGGCACTA GCAGCATTA TAACACCAAA  
 4990 5000 5010 5020 5030 5040  
 AAAGATAAAG CCACCTTTGC CTAGTCTTAC GAAACTGACA GAGGATAGAT GGAACAAGCC  
 5050 5060 5070 5080 5090 5100  
 CCAGAAGACC AAGGGCCACA GAGGGAGCCA CACAATCAAT GGACACTAGA GCTTTTAGAG  
 5110 5120 5130 5140 5150 5160  
 GAGCTTAAGA ATGAAGCTGT TAGACATTTT CCTAGGATTT GGCTCCATGG CTTAGGGCAA  
 5170 5180 5190 5200 5210 5220  
 CATATCTATG AAACCTTATGG GGATACTTGG GCAGGAGTGG AAGCCATAAT AAGAATTCTG  
 5230 5240 5250 5260 5270 5280  
 CAACAACCTGC TGTATTATCCA TTTCAGAATT GGGTGTGCGAC ATAGCAGAAT AGGCGTTACT  
 5290 5300 5310 5320 5330 5340  
 CAACAGAGGA GAGCAAGAAA TGGAGCCAGT AGATCCTAGA CTAGAGCCCT GGAAGCATCC  
 5350 5360 5370 5380 5390 5400  
 AGGAAGTCAG CCTAAACTG CTTGTACCAC TTGCTATTGT AAAAAGTGT GCTTTCATTG  
 5410 5420 5430 5440 5450 5460  
 CCAAGTTTGT TTCACAACAA AAGCCTTAGG CATCTCCTAT GGCAGGAAGA AGCGGAGACA  
 5470 5480 5490 5500 5510 5520  
 GCGACGAAGA CCTCCTCAAG GCAGTCAGAC TCATCAAGTT TCTCTATCAA AGCAGTAAGT  
 5530 5540 5550 5560 5570 5580  
 AGTACATGTA ATGCAACCTA TACAAATAGC AATAGCAGCA TTAGTAGTAG CAATAATAAT  
 5590 5600 5610 5620 5630 5640  
 AGCAATAGTT GTGTGGTCCA TAGTAATCAT AGAATATAGG AAAATATTAA GACAAAGAAA  
 5650 5660 5670 5680 5690 5700  
 AATAGACAGG TTAATTGATA GACTAATAGA AAGAGCAGAA GACAGTGGCA ATCAGAGTCA  
 5710 5720 5730 5740 5750 5760  
 AGGACAAATA TCAGCACTTG TGGAGATGGG GGTGGAATG GGGCACCATG CTCCTTGGGA  
 5770 5780 5790 5800 5810 5820  
 TATTGATGAT CTGTAGTGCT ACAGAAAAAT TGTGGGTCAC AGTCTATTAT GGGGTACCTG  
 5830 5840 5850 5860 5870 5880  
 TGTGGAAGGA AGCAACCACC ACTCTATTTT GTGCATCAGA TGCTAAAGCA TATGATACAG  
 5890 5900 5910 5920 5930 5940  
 AGGTACATAA TGTTTGGGCC ACACATGCCT GTGTACCCAC AGACCCCAAC CCACAAGAAG  
 5950 5960 5970 5980 5990 6000  
 TAGTATTGGT AAATGTGACA GAAAATTTTA ACATGTGGAA AAATGACATG GTAGAACAGA  
 6010 6020 6030 6040 6050 6060  
 TGCATGAGGA TATAATCAGT TTATGGGATC AAAGCCTAAA GCCATGTGTA AAATTAACCC  
 6070 6080 6090 6100 6110 6120  
 CACTCTGTGT TAGTTTAAAG TGCCTGATT TGGGGAATGC TACTAATACC AATAGTAGTA  
 6130 6140 6150 6160 6170 6180

10 NOV 04 - 23088

ATACCAATAG TAC GGG GAAATGATCA TGGAGAAA( AGATAAAA AACTGCTCTT

D F A

6170 6200 6210 6220 6230 6240

TCAATATCAG CACAAGCATA AGAGGTAAGG TGCAGAAAAG ATATGCATT TTTTATAAAC

6250 6260 6270 6280 6290 6300

TTGATATAAT ACCAATAGAT AATGATACTA CCAGCTATAC GTTGACAAGT TGTAACACCT

6310 6320 6330 6340 6350 6360

CAGTCATTAC ACAGGCCTGT CCAAAGGTAT CCTTTGAGCC AATTCCCATA CATTATTGCT

6370 6380 6390 6400 6410 6420

CCCCGGCTGG TTTTGGGATT CTAAAATGTA ATAATAAGAC GTTCAATGCA ACAGGACCAT

6430 6440 6450 6460 6470 6480

GTACAAATGT CAGCACAGTA CAATGTACAC ATGGAATTAG GCCAGTAGTA TCAACTCAAC

6490 6500 6510 6520 6530 6540

TGCTGTTGAA TGGCAGTCTA GCAGAAGAAG AGGTAGTAAT TAGATCTGCC AATTTACAG

6550 6560 6570 6580 6590 6600

ACAATGCTAA AACCATAATA GTACAGCTGA ACCAATCTGT AGAAATTAAT TGTACAAGAC

6610 6620 6630 6640 6650 6660

CCAACAACAA TACAAGAAAA AGTATCCGTA TCCAGAGGGG ACCAGGGAGA GCATTTGTTA

6670 6680 6690 6700 6710 6720

CAATAGGAAA AATAGGAAAT ATGAGACAAG CACATTGTA CATTAGTAGA GCAAAATGGA

6730 6740 6750 6760 6770 6780

ATGCCACTTT AAAACAGATA GCTAGCAAAT TAAGAGAACA ATTTGGAAAT AATAAAACAA

6790 6800 6810 6820 6830 6840

TAATCTTTAA GCAATCCTCA GGAGGGGACC CAGAAATTGT AACGCACAGT TTTAATTGCT

6850 6860 6870 6880 6890 6900

GAGGGGAATT TTTCTACTGT AATTCAACAC AACTGTTTAA TAGTACTTGG TTTAATAGTA

6910 6920 6930 6940 6950 6960

CTTGGAGTAC TGAAGGGTCA AATAACACTG AAGGAAGTGA CACAATCACA CTCCCATGCA

6970 6980 6990 7000 7010 7020

GAATAAAACA ATTTATAAAC ATGTGGCAGG AAGTAGGAAA AGCAATGTAT GCCCTCCCA

7030 7040 7050 7060 7070 7080

TCAGCGGACA AATTAGATGT TCATCAAATA TTACAGGGCT GCTATTAACA AGAGATGCTG

7090 7100 7110 7120 7130 7140

GTAATAACAA CAATGGGTCC GAGATCTTCA GACCTGGAGG AGGAGATATC AGGGACAATT

7150 7160 7170 7180 7190 7200

GGAGAAGTGA ATTATATAAA TATAAAGTAG TAAAAATTGA ACCATTAGGA GTAGCACCCA

7210 7220 7230 7240 7250 7260

CCAAGGCAAA GAGAAGAGTG GTGCAGAGAG AAAAAAGAGC AGTGGGAATA GGAGCTTTGT

7270 7280 7290 7300 7310 7320

TCCTTGGGTT CTTGGGAGCA GCAGGAAGCA CTATGGGGCC ACGGTCAATG ACGCTGACGG

7330 7340 7350 7360 7370 7380

TACAGGCCAG ACAATTATTG TCTGGTATAG TGCAGCAGCA GAACAATTTG CTGAGGGGCTA

7390 7400 7410 7420 7430 7440

16 NOV. 84- 29099

AGGCGCA ACAUCATC. TGCACCTCA CAGTCTGGGG CAT. CAG CTCCAGGCAA  
D F A

7450 7460 7470 7480 7490 7500  
GAATCCTGGC TGTGGAAAGA TACCTAAAGG ATCAACAGCT CCTGGGGATT TGGGGTTGCT

7510 7520 7530 7540 7550 7560  
CTGGAAAACCT CATTTGCACC ACTGCTGTGC CTTGGAATGC TAGTTGGAGT AATAAATCTC

7570 7580 7590 7600 7610 7620  
TGGAACAGAT TTGGAATAAC ATGACCTGGA TGGAGTGGGA CAGAGAAATT AACAATTACA

7630 7640 7650 7660 7670 7680  
CAAGCTTAAT ACATTCCTTA ATTGAAGAAT CGCAAAACCA GCAAGAAAAG AATGAACAAG

7690 7700 7710 7720 7730 7740  
AATTATTGGA ATTAGATAAA TGGGCAAGTT TGTGGAATTG GTTTAACATA ACAAATTGGC

7750 7760 7770 7780 7790 7800  
TGTGGTATAT AAAAATATTG ATAATGATAG TAGGAGGCTT GGTAGGTTTA AGAATAGTTT

7810 7820 7830 7840 7850 7860  
TTGCTGTACT TTCTATAGTG AATACAGTTA GGCAGGGATA TTCACCATT ACGTTTCAGA

7870 7880 7890 7900 7910 7920  
CCCACCTCCC AACCCCGAGG GGACCCGACA GGCCCGAAGG AATAGAAGAA GAAGGTCCAG

7930 7940 7950 7960 7970 7980  
AGAGAGACAG AGACAGATCC ATTCGATTAG TGAACGGATC CTTAGCACTT ATCTGGGACG

7990 8000 8010 8020 8030 8040  
ATCTGCGGAG CCTTGTGCCT CTTACGCTAC CACCGCTTGA GAGACTTACT CTTGATTGTA

8050 8060 8070 8080 8090 8100  
ACGAGGATTG TGGAACTTCT GGGACGCAGG GGGTGGGAAG CCCTCAAATA TTGTTGGAAT

8110 8120 8130 8140 8150 8160  
CTCCTACAGT ATTGGAGTCA GGAAGTAAAG AATAGTGCTG TTAGCTTGCT CAATGCCACA

8170 8180 8190 8200 8210 8220  
GCCATAGCAG TAGCTGAGGG GACAGATAGG GTTATAGAAG TAGTACAAGG AGCTTGTA

8230 8240 8250 8260 8270 8280  
GCTATTGCGC ACATACCTAG AAGAATAAGA CAGGGCTTGG AAAGCATTTT GCTATAAGAT

8290 8300 8310 8320 8330 8340  
GGGTGGCAAG TGGTCAAAAA GTAGTGTGCT TGGATGGCCT ACTCTAAGGG AAAGAATGAG

8350 8360 8370 8380 8390 8400  
ACGAGCTGAG CCAGCAGCAG ATGGGGTGGG AGCAGCATCT CGAGACCTGG AAAAACATGG

8410 8420 8430 8440 8450 8460  
AGCAATCACA AGTAGCAATA CAGCAGCTAC CAATGCTGCT TGTGCCTGGC TAGAAGCACA

8470 8480 8490 8500 8510 8520  
AGAGGAGGAG GAGGTGGGTT TTCCAGTCAC ACCTCAGGTA CCTTTAAGAC CAATGACTTA

8530 8540 8550 8560 8570 8580  
CAAGGCAGCT GTAGATCTTA GCCACTTTT AAAAGAAAAG GGGGGACTGG AAGGGCTAAT

8590 8600 8610 8620 8630 8640  
TCACTCCCAA CGAAGACAAG ATATCCTTGA TCTGTGGATC TACCACACAC AAGGCTACTT

8650 8660 8670 8680 8690 8700

D.F.A.

CCCTGATTGG CAG=ACTACA CACCAGGGCC AGGGGTCAGA TATCCACTCA CCTTTGGATG

6/26

8710	8720	8730	8740	8750	8760
GTGCTACAAG	CTAGTACCAG	TIGAGCCAGA	TAAGGTAGAA	GAGGCCAATA	AAGGAGAGAA

8770	8780	8790	8800	8810	8820
CACCAGCTTG	TTACACCTCG	TGAGCCTGCA	TGGAATGGAT	GACCTGAGA	GAGAAGTCTT

8830	8840	8850	8860	8870	8880
AGAGTGGAGG	TTTGACAGCC	GCCTAGCATT	TCATCACGTG	GCCCGAGAGC	TGCATCCGGA

8890	8900	8910	8920	8930	8940
GTAATTCAAG	AACTGCTGAC	ATCGAGCTTG	CTACAAGGGA	TCCGCTG	GGGACTTTCC

8950	8960	8970	8980	8990	9000
AGGGAGGCCG	GGCTGGGCG	GAAGTGGGGA	GTGGCGAGCC	CTCAGATGCT	GCAATAAAG

9010	9020	9030	9040	9050	9060
AGCTGCTTTT	TGCCTGTACT	GGGTCTCTCT	GGTTAGACCA	GATTTGAGCC	TGGGAGCTCT

9070	9080	9090	9100	0	0
CTGGCTAACT	AGGGAACCCA	CTGCTTAAGC	CTCAATAAAG	CTT	

Fig 26